LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-664301

# LC Recommendations for Data Transfer Nodes

*C. M. Harr, J.W. Long, T.M. Heer*

**January 29, 2015**

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

# Hardware Recommendations for DTN nodes



## Introduction

These recommendations are in regards to the LLNL Data Transfer Nodes (DTNs), which are used to transfer Climate Storage System (CSS) data between nodes of the ESGF grid. The hardware specifications listed are to be considered a base configuration for a node and can be enhanced in a variety of ways to suit a particular site.

## Hardware Specifications

### Base

- 1U chassis with dual-socket "Socket R" motherboard
    - Eg: Intel Server System R1304GZ4GC with a Intel Server Board S2600GZ4
    - Don't forget rails
- 2x 750W power supplies
- 2x Intel Xeon E5-2667 v3 (Haswell) 2.6GHz, 8C CPU
- 64GB (8x8GB) DDR3 ECC memory
- 1x Intel/LSI RMS25CB040 RAID 0/1/5/6/10/50/60 4-port Mezzanine card
    - 8-port RMS25CB080 also available
- 1x Intel® RAID Smart Battery (AXXRSBBU9)
- 1x Intel® RAID Maintenance Free Backup (AXXRMFBU2)
- 2x 1TB SATA 3.5" hot-swappable HDDs
- 2x Intel 82559ES Dual-port 10Gb Ethernet NIC w/ site-appropriate SFPs.

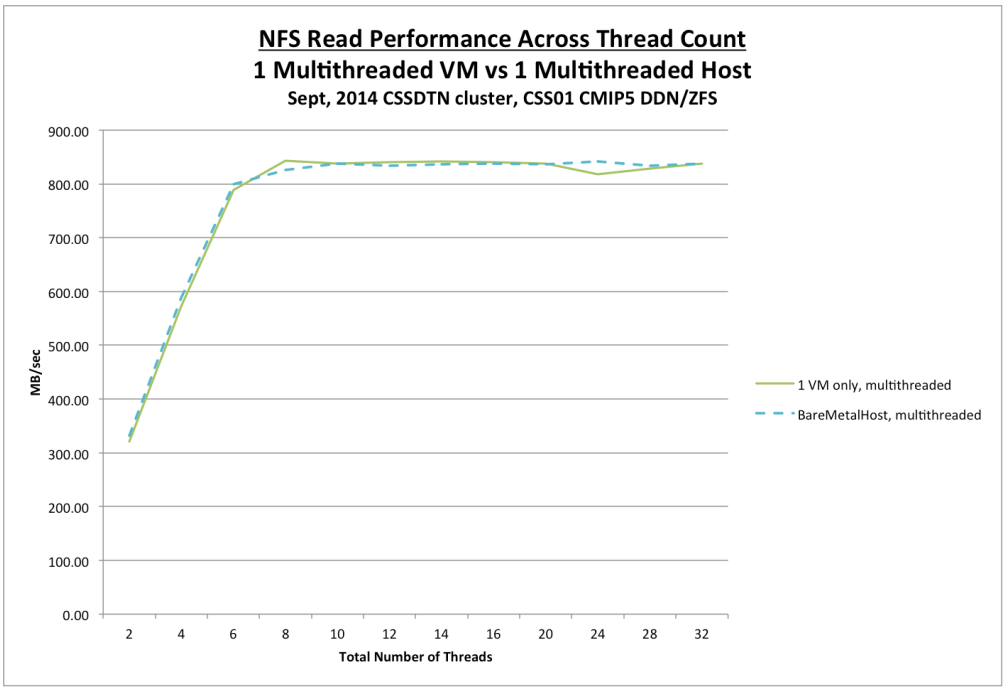
### Local Storage Option

The Base configuration assumes data is transferred directly to NFS mounts using the second 10Gb Ethernet interface. If a site prefers instead to have locally-attached storage to the DTN node, the following Fibre Channel adaptor is recommended:

- 1x Quad-port Qlogic FC8 HBA

## NFS Performance

With a configuration similar to the base configuration above, (but with 128GB memory), LLNL was able to achieve NFS read performance > 820 MB/s with 8+ threads (both across a VM and a bare-metal node).




## GridFTP Performance

Using the same alternate configuration used for the NFS tests above, disk-to-disk tests were done with basic GridFTP using the globus_url_copy command line tool. Transfers were initiated from a NERSC DTN system, and data were read from either the bare metal LLNL DTN host or a set of VMs on an equivalent host. The transfer tests involved copying a set of 1GB files containing random data. In the VM case, each transfer occurred simultaneously in a separate VM. In all cases, each individual file transfer used four threads to achieve more parallelism.

NOTE: There is a significant variability in the numbers because tests were done on production systems: the Climate Storage System at LLNL and the DTN nodes and scratch file systems are NERSC are all utilized 24 x 7 by users around the world.

*Bare Metal vs VMs*

Comparison of performance between equivalently configured VMs versus the same number of threads on a single host. Every attempt to access "cold" data was made to remove the effects of NFS caching, except as represented in the last column.

| | MB/s | | |
|---|---|---|---|
| **Num** | **VM** | **Bare Metal** | **Bare Metal **Cached Data**** |
| 1 | 207 | 221 | 485 |
| 2 | 279 | 276 | 787 |
| 3 | 330 | 393 | 877 |
| 4 | 422 | 437 | 917 |
| 5 | 448 | 495 | 977 |

As a result of these findings, LLNL DTNs are not based on VMs. The slight performance penalty plus the complexity of managing DTNs as a cluster of VMs more than offset the flexibility VMs offer.

## Future Looking

Future implementations of the DTN nodes may need to interface with a 40Gb Ethernet connection for state-of-the-art performance. In that use case, LLNL recommends the following configuration:

- Single-socket high-frequency CPU, e.g. the Intel Xeon E5-2637v3 (Haswell) 3.5GHz, 4-core CPU
    - LLNL has found a single-socket, high-frequency CPU provides superior performance with the 40GbE NIC
- Intel XL710 QSFP+ PCI-E x8 Single-port 40GbE NIC
    - Include Intel XL710 QSFP+ SR Optical Module